37359/JFO/B600

WHAT IS CLAIMED IS:

1. A method for reducing a precision of an input datum having precision portion and a loss portion, comprising:

a. comparing the loss portion to a preselected threshold value, f_t;

b. determining a selectable bias, α, responsive to the loss portion being in a defined relation to the preselected threshold value, f_t; and

c. combining the precision portion with α, creating a reduced precision datum thereby,

wherein α corresponds to a predetermined characteristic of one of α, the input datum, the reduced precision datum, and a combination thereof.

2. The method of claim 1, wherein determining the selectable bias further comprises one of:

a. assigning a first value to α, responsive to the loss portion being substantially equal to f_t;

b. assigning a second value to α, responsive to the loss portion being less than f_t; and

c. assigning a third value to α, responsive to the loss portion being greater than f_t.

3. The method of claim 1, further comprising determining the selectable bias responsive to a predetermined characteristic

of a plurality of input data relative to a corresponding plurality of reduced precision data.

4. The method of claim 1, further comprising determining the selectable bias responsive to a predetermined characteristic attributable to reducing the precision of the input datum.

5. The method of claim 1, further comprising determining the selectable bias responsive to the predetermined characteristic of the selectable bias, the predetermined characteristic being the mean value of a plurality of selectable bias values.

6. The method of claim 2, further comprising determining the selectable bias responsive to a predetermined characteristic of a plurality of input data relative to a corresponding plurality of reduced precision data, and the predetermined characteristic being attributable to reducing the precision.

7. The method of claim 6, wherein the predetermined characteristic is a predetermined mean error value.

8. The method of claim 2, further comprising determining the selectable bias responsive to a predetermined characteristic

of one of input data, a corresponding reduced precision data, and a combination thereof.

9. The method of claim 8, wherein the predetermined characteristic comprises a predetermined statistical value.

10. The method of claim 4, wherein the predetermined characteristic comprises a predetermined mean error value of the plurality of reduced precision data relative to a corresponding plurality of input data.

11. The method of claim 9, wherein the predetermined statistical value comprises the mean value of the reduced precision data relative to a corresponding plurality of finite-precision fixed point input data.

12. The method of claim 2, further comprising assigning a fourth value to α, responsive to a being substantially equal to f_t, the fourth value being in a predefined relationship with the first value.

13. The method of claim 12, further comprising determining the selectable bias responsive to a predetermined characteristic of input data relative to corresponding reduced precision

data, and the predetermined characteristic being a preselected mean error value associated therewith.

14. The method of claim 12, wherein:

a. the f_t is approximately equal to 0.5_{10};

b. the first value is 1 when the value of the loss portion substantially equals about 0.5_{10}, the input datum is a negative-valued datum, with the first value being added to the precision portion;

c. the second value is zero when value of the loss portion is less than about 0.5_{10};

d. the third value is 1 when the value of the loss portion is greater than about 0.5_{10}, with the third value being added to the precision portion;

e. the fourth value is 0 when the loss portion substantially equals about 0.5_{10}, and the input datum is a positive-valued datum; and

f. the preselected mean error value relative to the input datum and the reduced precision datum is minimized.

15. The method of claim 11, wherein:

a. f_t is substantially equal to 0.5_{10};

b. the first value is a current first value being selected to be one of '1' and '0' when the value of

the loss portion substantially equals about 0.5_{10}, in a predefined relationship to a previous first value;

c. the second value is zero when the loss portion is less than about 0.5_{10}; and

d. the third value is 1 when the loss portion is greater than about 0.5_{10}, with the third value is added to the value of the precision portion.

16. The method of claim 14, wherein the predefined relationship is an alternating relationship.

17. The method of claim 16, wherein the alternating relationship is a toggle relationship with the current first value being zero if the previous first value was 1, and the current first value being 1 if the previous first value was zero, and wherein the preselected mean error value is minimized responsive to the alternating relationship.

18. The method of claim 15, wherein the alternating relationship includes a selectable number of 1's being interleaved with a selectable number of zeros, the mean value of the reduced precision data being responsive to the alternating relationship.

19. The method of claim 2, wherein each of the input datum and the reduced precision datum are represented by two's complement fixed point values.

20. The method of claim 16, wherein the alternating relationship includes a selected pseudorandom sequence of data bits.

21. A method for rounding a first datum, $\boldsymbol{X}$, having precision of $\boldsymbol{a}$ digits, to a second datum, $\hat{\boldsymbol{X}}$, having precision of $\boldsymbol{b}$ digits, wherein $\boldsymbol{a} > \boldsymbol{b}$, first $\boldsymbol{b}$ digits of $\boldsymbol{X}$ being a precision portion, and remaining $\boldsymbol{a-b}$ digits of $\boldsymbol{X}$ being a loss portion, the method comprising:

a. evaluating the loss portion relative to a preselected rounding threshold value;

b. if the loss portion is substantially equal to the preselected threshold, then defining $\hat{\boldsymbol{X}}$ according to the equation:

$$\hat{\boldsymbol{X}} = \boldsymbol{X} + \mathbf{2}^{-(b+1)}\alpha,$$

where α is a selectable bias represented by a rounding digit;

c. if the loss portion is not substantially equal to the preselected threshold, then defining $\hat{\boldsymbol{X}}$ according to the equation:

$\hat{\boldsymbol{X}} = \boldsymbol{X} + \mathbf{2}^{-(b+1)}$; and

d. eliminating the loss portion of $\boldsymbol{X}$, producing $\hat{\boldsymbol{X}}$ thereby.

22. The method of claim 21, wherein selectable bias α is representative of a predetermined characteristic of one of $\boldsymbol{X}$, $\hat{\boldsymbol{X}}$, α, and a combination thereof.

23. The method of claim 22, wherein the preselected threshold is substantially equivalent to 0.5_{10}.

24. The method of claim 23, wherein the predetermined characteristic comprises a preselected mean error value of $\hat{\boldsymbol{X}}$ relative to $\boldsymbol{X}$.

25. The method of claim 24, wherein the preselected mean error value, $\boldsymbol{E(e)}$, is substantially defined by the equation:

$$\boldsymbol{E(e) = 2^{-a}(E(\alpha) - \tfrac{1}{2})},$$

where $\boldsymbol{E(\alpha)}$ is a mean value of selectable bias α.

26. The method of claim 25 wherein the mean value of the selectable bias is substantially within the range of

$$\boldsymbol{0.0 \leq E(\alpha) \prec 1.0}$$

27. The method of claim 26, wherein the mean value of the selectable bias, $\boldsymbol{E(\alpha)}$, is approximately equal to

preselected mean error value, $\boldsymbol{E(e)}$, and $\boldsymbol{E(\alpha)}$ is approximately zero.

28. The method of claim 27, wherein the predetermined characteristic further comprises a preselected error variance value, σ_e^2, substantially defined by the equation:

$$\sigma_e^2 = \frac{2^{-2b} + 2^{-(2a-1)}}{12}$$

29. The method of claim 28, wherein the rounding digit is selected from a alternating sequence of digits in the pair of digits <0,1>.

30 The method of claim 28, wherein the rounding digit is selected from a pseudorandom sequence of binary digits.

31. A method for rounding a first two's complement fixed point datum, $\boldsymbol{X}$, having an integer part of $\boldsymbol{n}$ bits, a fractional part of $\boldsymbol{a}$ bits the integer part, and sign bit, $\boldsymbol{s_i}$, to a second two's complement fixed point datum, $\boldsymbol{\hat{X}}$, having a fractional part of $\boldsymbol{b}$ bits following the radix point, where $\boldsymbol{a}$ and $\boldsymbol{b}$ are representative of the respective precisions of $\boldsymbol{X}$ and $\boldsymbol{\hat{X}}$, and where $\boldsymbol{a} > \boldsymbol{b}$, comprising:

a. evaluating the fractional part of X and defining y as the most significant bit (MSB) of the ***a*** bits;

b. if the first bit following the radix point of X is equal to a 1 bit trailed by $(a-1)$ zero bits, then defining $\hat{X}$ according to the equation:

$$\hat{X} = n + s_i$$

and

c. otherwise, defining $\hat{X}$ according to the equation:

$$\hat{X} = n + y$$

32. The method of claim 31, wherein the occurrence of positive numbers and negative numbers in a plurality of the datum, X, is substantially equiprobable.

33. A method for rounding signal values, comprising:

a. detecting a predetermined state value wherein rounding is desired; and

b. rounding the state value according to one of

i. an alternating round-up/round-down method and

ii. a sign addition round-up/round-down method.

34. An arithmetic device, comprising a bias generator producing a selectable bias α, responsive to a predetermined signal characteristic, the device receiving an input signal and

coupling the selectable bias α thereto.

35. The arithmetic device of claim 34, further comprising a combiner coupled to the bias generator, the combiner receiving and combining the input signal and the selectable bias α, and producing an output signal.

36. The arithmetic device of claim 34 further comprising wherein the bias generator further comprises a comparator for comparing the input signal to a preselected threshold value, the comparator urging the bias generator to produce the selectable bias α responsive to the preselected threshold value.